Weekly Report

December 11, 2016

1 Work

Line Field Based Trajectory Analysis

I generate trajectory data in text formate and send it to Ding Ziang.

Semantic Trajectory

This week, I read a paper Discovering Urban Functional Zones Using Latent Activity Trajectories by Zheng Yu. In this paper, authors developed a topic-modeling- based approach (DMR topic model)to cluster the segmented regions into functional zones leveraging mobility(flow) and location semantics(poi). I think it is more reasonable than author topic model, because author topic model mainly focus on poi distribution. Then, I found a project called MALLET which consist of LDA and DMR topic model. I download some Chinese documents form sougou lab and make Chinese word segmentation using python library called jieba. The MALLET works fine with LDA. However, I get error message when I estimate topic with DMR topic model.

2 Plan for next week

- Read some papers about topic model.

3 Paper Reading

3.1 Discovering regions of different functions in a city using human mobility and pois

这篇文章主要使用DMR topic model来探索城市的功能区，其中，结合了出租车和刷卡记录等轨迹数据，以及北京地区的poi数据。

3.2 Discovering Urban Functional Zones Using Latent Activity Trajectories

这篇文章是上一篇的扩展版，增加了一些城市根据路网划分，协同过滤后的poi分布。

3.3 Surprise！Bayesian Weighting for De-Biasing Thematic Maps

主题地图经常用来编码数据在地图上的分布，然而不论是使用绝对数值还是比例，都会受到整体人口的影响。人口多，一般绝对数值就比较多；人口少，因为样本量少，有一些比例值就会比较高。所以作者使用贝叶斯权衡的方法，计算一个语气和实际数据中得到的一个surprise值进行编码。

3.4 VisFlow – Web-based Visualization Framework for Tabular Data with a Subset Flow Model

本文提出了VisFlow框架，支持用户以可视的方式直接自定义数据分析流程从而帮助用户的分析，克服了过去计算性数据流系统的交互问题，支持在数据流中进行交互选择，brushing和linking。

3.5 Topic Models Conditioned on Arbitrary Features with Dirichlet-multinomial Regression

这篇文章是DMR topic model的原始论文，从中我惊讶的发现论文中也使用了MALLET这个工具。看了作者之后我发现，本文的作者和MALLET作者是一样的。于是，我认为MALLET中的DMR主题模型应该是没问题的，大概是我的打开方式有问题。